DOCUMENT RESUME

ED 383 773 TM 023 855

AUTHOR Tushnet, Naida C.
TITLE Toward a General Approach to Multisite Program Evaluation.
PUB DATE Apr 95
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Context Effect; Distance Education; Elementary Secondary Education; *Evaluation Methods; Evaluation Problems; Evaluation Utilization; *Federal Programs; Partnerships in Education; Program Development; *Program Evaluation; Program Implementation; *Values
IDENTIFIERS *Deconstruction; Educational Partnerships Program (OERI); *Multiple Site Studies; Star Schools; Theory Based Evaluation

ABSTRACT

Evaluations of multisite programs present special challenges to evaluators concerned with the use of evaluation results. This paper offers an approach that holds promise for mitigating the problems inherent in evaluating programs at multiple sites. A brief analysis is presented of the challenges these programs present and the reasons these challenges exist. In sum, these challenges exist because each site implements a program differently. Two federal education programs, the Educational Partnerships Program and the Star Schools distance learning program, are described and are used as examples throughout the paper. A second section describes a general theoretical approach to evaluation that is based on the theory-driven evaluation approach of H.-T. Chen (1989) and the work of Jacques Derrida (1981) and others who emphasize the deconstruction of concepts and activities. The evaluation framework, illustrated in figure 1, incorporates the elements of: (1) context; (2) problem definition; (3) program; (4) funded projects; (5) activities; and (6) outcomes. Addressing program developers' values and assumptions is particularly useful in multisite program evaluations. (Contains 42 references.) (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 383 773

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. Tushnet

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

# Toward a General Approach to Multisite Program Evaluation

Naida C. Tushnet
Southwest Regional Laboratory

American Educational Research Association Annual Meeting
San Francisco
April, 1995

TM023855



BEST COPY AVAILABLE

# INTRODUCTION

Evaluations of multisite programs present special challenges to evaluators concerned with the use of evaluation results (Tushnet, 1993). This paper offers an approach that holds promise for mitigating the problems inherent in evaluating programs implemented in a number of different sites. The Introduction of this paper includes a brief analysis of the challenges presented by multisite program evaluations and the reasons these challenges exist. In addition, the Introduction briefly describes two federal education programs, the Educational Partnerships Program and the Star Schools distance learning program, because they will be used as examples throughout the paper. The second section of the paper describes the general theoretical approach to evaluation.

## Challenges of Multisite Evaluations

Most federal programs, and frequently state and foundation-supported programs as well, are implemented at several sites. Program intentions differ, depending on the nature of the program, with some federal programs designed to be similar across sites and others in which variation is both expected and desired. For example, Chapter 1 of the Elementary and Secondary Education Act provides educational support for educationally disadvantaged children across the nation. Although different treatments are expected depending on the characteristics of students and communities, presumably there are fairly large similarities across sites. In contrast, demonstration programs, such as the Educational Partnerships Program described in this paper, are designed to have maximum variability. In the middle are programs, such as Star Schools, that fund several projects expected to differ from one another; however, activities within sites served by a single project are expected to be similar.

Whatever the design of the program, since the 1970s, researchers and evaluators have noted that program implementation differs at each site (Berman & McLaughlin, 1976). Goggin, Bowman, Lester, and O'Toole (1990) note that even for laws, implementation varies according to context. Variations are even greater for programs. This variation across sites leads to one challenge to evaluators because it may be difficult to untangle program effects from effects generated by the conditions at a single site, including site biases about program participants (Sinacore & Turpin, 1991).

A second challenge comes from the reasons implementations vary by site. Some variation may result from a local project's assessment of the availability of appropriate staff. Local leaders may, for example, change the configuration of project services because individuals with the requisite background and experience are unavailable, as when a program that includes psychological support for underachieving students uses support groups rather than individual counseling because trained social workers or psychologists are not available in the community. As

Mowbray and Herman (1991) put it, "How does one assess the effect that variability in staff composition has on client outcomes?" (p. 49).

Similarly, some variation in local implementation may result from differences in how personnel at different sites assess client needs. Within Chapter 1, for example, two types of treatments are allowed by regulation. One focuses on individual students, providing remediation and other supports for students identified as not achieving as well as their age mates. The other focuses on the school. These schoolwide Chapter 1 projects are allowed when the number of disadvantaged children is great enough that local educators believe that across-the-school changes are needed to provide adequate educational opportunities.

Some variation is difficult to predict and hard to measure. For example, in the Star Schools Program, students taking the same course transmitted by satellite to multiple sites had very different experiences. The national evaluation revealed that the differences were not due to "objective" characteristics of the students, the schools, the studio teacher, the content of the course, or even the background and education of the local facilitator. Less tangible matters, such as the interest of the facilitator and school and district leaders, seemed to account for the differences (Tushnet et al., 1994).

Perhaps most important, individuals at different sites interpret programs differently. According to Goggin et al. (1990), "Interpretation is a function of context...a single message, such as a federal statute, may be interpreted differently in different states" (p. 33). Fullan (with Steigelbauer, 1991) notes that educational innovation involves continuous, and developing, understanding of the substance of the change by school personnel. As a result, program implementation is an ongoing process, during which the nature of the innovation as well as current practice may change (Berman & McLaughlin, 1977).

Multisite evaluations also are challenging because they most frequently rely on some amount of decentralized data collection (Sinacore & Turpin, 1991). To the extent that program developers allow local sites to determine the data to be collected, there will be variation in what information is available to the cross-site program evaluator. Further, in both the Educational Partnerships and Star Schools programs, local sites were found to collect little information about students or other service recipients. In particular, little information was available related to demographics and outcomes. Even when local sites collect data, problems of quality control arise (Sinacore & Turpin, 1991).

In sum, multisite evaluations are challenging because each site implements a program somewhat differently. As a result, what is assumed to be a single program comprises, in reality, multiple treatments that may be impossible to describe in advance. Consequently, program evaluators should be as interested in describing the program, in its many manifestations, as in assessing effects. As will be argued later in this paper, descriptions of programs, projects, and activities are data for deconstructing assumptions and values and the basis for the usefulness of the evaluation.

## Descriptions of the Programs

Among the programs being evaluated by the Policy and Evaluation Center of the Southwest Regional Laboratory (SWRL) are two national, multisite programs. The first is a demonstration program, the Educational Partnerships Program, designed to provide information about the role of partnerships that involve public schools, businesses, social service agencies, and cultural institutions in schools and their support in the community. The second is the Star Schools Program, a federally supported distance education program designed to improve opportunities for students in remote rural and educationally disadvantaged urban schools. Each of these programs is described in turn.

### *The Educational Partnerships Program*

Congress enacted the Educational Partnerships Act in 1988 (title VI, subtitle A, Chapter 5 of the Omnibus Trade and Competitiveness Act of 1988, Pub. L. 100-418). The intent of the legislation was to stimulate the creation of educational partnerships in order to demonstrate their contributions to educational reform. The authorizing legislation requires documenting the partnerships that receive assistance, assessing their impact on educational institutions, evaluating the extent to which they improve their communities' climate for support of education, and identifying promising activities. The legislation goes beyond the simple evaluation question of whether funded projects meet their own objectives. It questions program impact on the participating institutions and their communities. The broader question is whether partnerships (and the particular activities they sponsor) can be a force to renew education and encourage community support.

The Educational Partnerships Program is administered by the Educational Networks Division, Programs for the Improvement of Practice, Office of Educational Research and Improvement (OERI), U.S. Department of Education. OERI funded four cycles of partnership projects for a total of 29 projects. The projects differ in structure and objectives. They include partnerships designed to facilitate the transition from school to work; improve instruction in mathematics and science; provide opportunities for gifted students, those at risk of school failure, and noncollege bound students; and stimulate systemic reform. The partnerships include a variety of configurations of businesses and industries, cultural institutions, health and human service agencies, institutions of higher education, state education agencies, and elementary and secondary schools.

The documentation and evaluation study is being conducted during a three-year period. The study was designed to identify partnership structures and activities associated with successful implementation, positive impact, and long term institutionalization. In addition to sponsoring a program evaluation, OERI required each funded project to implement a local evaluation. SWRL

was charged with providing technical assistance to the local evaluators. OERI believed that the technical assistance would result in higher quality local evaluations.

### *The Star Schools Program*

The federal government, through what is known as the Star Schools Program Assistance Act (20 U.S.C. 4081), has encouraged:

> Improved instruction in mathematics, science, foreign languages, and other subjects, such as literacy skills and vocational education...to serve underserved populations, including disadvantaged, illiterate, limited-English proficient, and disabled students through distance learning technologies. (PL 102-103 Aug. 17, 1991)

The Star Schools Program supports a variety of services, including:

- developing or acquiring programming in various curriculum areas;
- supporting teacher training and staff development to improve instruction in mathematics, science, and foreign languages;
- providing tutorial services for students by using a variety of technologies;
- maintaining testing services for courses offered; and
- supporting the acquisition of telecommunication facilities and equipment, both by projects and by participating schools.

The program evaluation focuses on a series of questions posed in the authorization of Star Schools. The questions address issues related to project organization and impact on students, staff, and schools. Congress expressed interest in: (a) whether courses generated through Star Schools last beyond federal funding, (b) the extent to which Star Schools increased access to courses for students, and (c) the effectiveness of Star Schools courses and staff development opportunities.

# THEORETICAL APPROACH

The framework for creating useful multisite evaluations draws greatly from two sources. First, it reflects some root concerns addressed by Chen (1989) in his work on "theory-driven evaluation," although it lacks the heavy quantitative framework. The second source is the work of Derrida (1981) and others who direct attention to "deconstruction" of concepts and activities, although it lacks the heavy political ideology generally associated with the deconstructionists. In this section, I discuss the contributions of each as well as the reasons their ideas are promising for increasing the usefulness of evaluations. Examples from the two programs are interwoven with the theoretical grounding. The theoretical discussions will be followed by an explanation of each element of the overall evaluation approach as well as descriptions of appropriate methods to use in explicating and deconstructing program assumptions and values.

## Theory-Driven Evaluation

Chen and his colleagues argue that an overriding purpose of evaluation is to interpret the effects of various treatments on the populations of concern. Doing so, they argue, requires going beyond the "black box" approach (Lipsey, 1993) that simply looks at outcomes to understanding what happened to produce the results, at least in nonexperimental designs. Because it is difficult and generally undesirable to create true experiments for social programs, alternative approaches to understanding program dynamics are needed. Proponents of theory-driven evaluation maintain that "treatment theory" should form the basis of the evaluation. Lipsey (1993) lists the following elements of adequate treatment theory:

> (1) a problem definition that specifies what condition is treatable, for which populations, and under what circumstances...;(2) specification of the critical inputs, ...the interrelationships among the inputs, and the basis for judging magnitude or strength of inputs...; (3) the important steps, links, phases, or parameters of the transformation process that the treatment brings about, the intervening or mediating variables on which the process is contingent, and the crucial interactions with individual differences, timing, mode of delivery, or other relevant circumstances; and (4) specification of the expected output...and the interrelationships or contingencies among the outputs. (p. 11)

The evaluation is driven by the:

> theory or model of how a program is supposed to work. The theory of the program's operation is then used as a basis for designing an evaluation of the program. By explicitly relating program processes to outcomes, this approach seeks to avoid the limitations of the methods-driven and black-box approaches to evaluation. (Smith, 1993, p. 3)

The treatment theory, then, provides the evaluator with the framework that governs the evaluation. It not only generates appropriate questions about outcomes, but also directs attention to

questions about the processes that lead to the outcomes. Evaluators can seek indicators about the populations receiving treatment, the inputs, the various necessary linkages, and outcomes.

Frequently, as in the case of the Educational Partnerships Program, program developers have only implicit theories about how the program is supposed to work. One desired outcome was increased community support for education. The logical chain from the formation of partnerships to increased community support was never spelled out, but rather was taken as a matter of faith. As evaluators, we could look for indications of greater community support (e.g., passage of bond and tax levies, positive newspaper reports, responses to interview questions), but interpretation was difficult unless we could construct a logic (a "treatment theory") of how partnerships engender such positive outcomes. The logical line was as follows:

> In many communities, fewer adults have children attending public school than in earlier times. Consequently, adult views about schools are based on impressions gathered from newspaper articles about youth and casual interaction with young people in such places as shopping malls. Neither of these sources leads to a positive image for schools. Educational partnerships, on the other hand, provide an opportunity for adults to get to know young people as well as teachers and other school personnel. Such knowledge will create understanding for the challenges faced by schools and the honest, hardworking efforts school people make to meet those challenges. As a result, participants in educational partnerships will become advocates for schools among their friends, thus increasing support for education.

The usefulness of presenting this logical construction to program developers was manifested in two ways. First, on the federal level, the developers were able to judge whether particular projects' approaches to a partnership had a plausible relationship to this outcome. According to Chen and Rossi (1983), a lot of approaches are implausible. In this light, it is not surprising that few of the partnerships had an impact on community support because most involved few community members. It is unrealistic to expect that such a goal will be met simply as a result of an educational partnership. Federal personnel as well as project staff members began to look at how to use the partnerships to increase community awareness rather than rely on their existence to do so. This led to some altered project designs.

The second reason for making the program theory explicit relates to the multisite nature of the evaluations. By raising to consciousness the implicit theories held at both the federal and project sites, differences could be noted. This was most evident in the Evaluation of the Star Schools Program. The Star Schools legislation and program announcements focused on three distinct goals: increasing access to courses; demonstrating educationally sound uses of technology; and supporting reform, particularly in schools serving disadvantaged children, through distance learning. For federal policymakers and program developers, the three goals were envisioned as mutually supportive, but funded projects tended to emphasize one goal. Each project had an implicit, or more rarely, explicit, program theory guiding its activities. For example, projects that focused on providing courses to students, mainly in rural areas, to which they would otherwise not have access built few support mechanisms for the facilitators at the receiving site, identified

teachers with excellent qualifications to serve as studio teachers, and took full responsibility for the quality of the curriculum and instruction. The logic was:

> The problem is that some students don't have the opportunity to take some courses because their schools are too small to offer them or there are too few students interested to warrant hiring a teacher. All that is needed is a good teacher in the studio teaching a course that other students receive. At the site, the facilitator will simply make sure students follow the curriculum, do their homework, and take their tests. Student learning of fairly traditional subject matter is the expected effect of distance learning.

In contrast, the projects that emphasized distance learning as an agent of educational reform had the following theory:

> Schooling as currently experienced does not meet the needs of learners or society. Major changes are needed to support a new paradigm of education, and distance learning can provide impetus for those changes. Studio teachers teach in a new "constructivist" manner, both to provide students with high quality opportunities for learning and to model cutting edge pedagogy for receiving site teachers. The studio teacher and the classroom teacher are a team, and effects should be on both students and teachers.

The situation resulted in virtually all project staff members feeling misunderstood and devalued. They believed that they were always being compared to the other projects and always found lacking. Each believed that the standards by which they were judged were designed for a different type of project. As a result, it was difficult for federal program personnel to create the collaboration they desired among grantees. The evaluation report's analysis of these underlying, different, program assumptions allowed projects to begin more productive conversations. (However, all projects believed the report underplayed the importance of their particular approach!)

The examples given so far are those in which the program theory was implicit. In fact, implicit theories were more common in both the Star Schools and Educational Partnerships programs. However, some projects, particularly partnerships concerned with the transition from school to work, had explicitly stated and developed program theories. In such cases, the work of the evaluator was easier because the challenge was limited to developing appropriate indicators for the elements in the theory rather than first explicating the theory and then developing the indicators. A further complication arose for project staff, both in the partnerships and Star Schools programs, who revised and refined their theories—and actions—as a result of learning from the evaluation or experience. While such strategic adjustments were more likely to lead to positive outcomes than maintaining a theory and its related activities (Tushnet et al., 1993), the changes also require continuous adjustments by the evaluators. Program staff use of evaluation results seems to lead to harder, although more satisfying, work for the evaluator.

## Deconstruction

The contribution of deconstructivism to multisite program evaluation lies in its focus on "problematizing" that which is taken for granted. Deconstructionists approach texts in literature or data in research with an "all-encompassing critical skepticism about knowledge" (Vidich & Lyman, 1994, p. 41). Such skepticism frames evaluation questions of how programs should operate and who they should benefit. It also directs the evaluator to look for hidden messages embedded in the program and its implementation, particularly those that reinforce social and economic inequalities. The evaluator deconstructs in order to gain knowledge of the real social justice issues and outcomes of a program. The evaluator also deconstructs her or his own influences on design, data collection, and interpretation (Derrida, 1976) and sometimes offers alternative perspectives on the same piece of evidence.

Although different in emphasis, deconstructivism relates closely to constructivist epistemologies (Schwandt, 1994). Both are concerned with how individuals interpret the meaning of their position and actions, and both focus on meaning as constructed through social interaction (Berger & Luckmann, 1966). Because meaning is constructed, according to this argument, it reflects commonly accepted relationships and has a tendency to support the status quo. It is the task of *de*construction, then, to raise questions about relationships and ideas—to make them problematic. Of prime importance to deconstructionists are power relationships. For example, in looking at the large numbers of female teachers and small numbers of female administrators, the deconstructionist asks questions about the nature of power and gender in schools and society. Even more radically, deconstructionists "expose the ideological and political meanings that circulate within the text, particularly those which hide or displace racial, class, ethnic and gender biases" (Denzin, 1992, p. 151). For evaluators, the issue is not so much "text" as "program." Or as Flax (1987) puts it, deconstructivism seeks "to distance us from and make us skeptical about beliefs concerning truth, knowledge, power, the self and language that are often taken from granted within and serve as legitimation for Western culture" (p. 624).

Among the issues deconstructionism raises for evaluation is the question of "how voices of participants are to be heard, with what authority, and in what form" (Olesen, 1994, p. 167). Olesen argues that the question of voice is closely related to "the account," that is, the narrative that frames the data. She also notes that typical interviews and questionnaires, even open-ended interviews, of "participants" are frequently insufficient to give voice to those who have been voiceless. The questions frame the issues—help construct the reality—as does the age, gender, ethnicity, and class of the interviewer.

Clearly, the deconstructionist project has political implications, but deconstructionists argue that all research has (disguised) political implications. One job of the evaluator, then, is to probe for the political meanings of programs. For multisite programs, the political meanings are constructed at the program (national) level and within each site. Further, the deconstructive

program creates the appropriate skepticism about responses to questions, and, indeed, the questions themselves.

An example of how a deconstructive cast of mind can advance an evaluation is drawn from the Star Schools Program. As noted above, the program had three rather different types of goals, and each project seemed to chose one for focus, although no formal decisions were made to that effect. One common thread ran across all program goals, and that was concern for educationally disadvantaged children and youth, with authorization and appropriation tied to Chapter 1. Indeed, all projects involved students who were disadvantaged, either by reason of socioeconomic conditions or living in rural, isolated areas. However, the data indicated great differences in the types of programs offered to youth in schools serving low economic areas and those in rural schools:

> The data related to student characteristics in schools served indicate differences between schools receiving full courses and those receiving supplementary instruction. Students enrolled in advanced courses and foreign languages are more likely to be white and less likely to be poor than are students enrolled in supplemental courses. When such data are reported within a single school, observers frequently note that tracking is occurring. (Tushnet et al., 1994, p. 97)

In this example, the evaluators made definitions-in-use of "educationally disadvantaged" problematic. Rather than merely reporting different types of disadvantage addressed by the distance learning program, we looked at the specific characteristics by type of course. In this case, we found that the multiple goals of the program coupled with each project's different kind of attention to educationally disadvantaged students had an unintended twist.

Deconstructing program theories, particularly when they are implicit theories, is central to the evaluation approach. As the example indicates, even within a single program, local implementers frequently differ in their understanding of the problem and what will solve it. The following contrasts the theories held by one project that provided students with access to courses, mainly in rural areas, and another that embraced supplementary instruction in the cause of educational reform:

1. Students, mainly in rural areas but occasionally in urban centers, are deprived of equal educational opportunity because the schools they attend are not able to offer particular courses, mainly in science and foreign languages. Consequently, offering such courses through distance learning will mitigate their disadvantage.

2. In urban centers, students do not have the same educational opportunities as their suburban peers because urban teachers are less likely to have participated in staff development on new approaches to teaching mathematics. Further, because students enter school with less home and community experience in mathematics, they need supplementary instruction if they are to achieve what they are capable of achieving. Both problems lead to offering supplemental instruction in which the studio teacher models appropriate behavior.

The deconstructionist notes the difference in defining "disadvantage," and notes as well that the first approach is likely to reinforce current distributions of power by increasing the number of middle class whites who are eligible for elite institutions. Whether the second approach supports or changes social relations requires additional information, but it has the potential for making changes.

Perhaps the most significant example of how deconstruction assisted in a program evaluation comes from the study of the Educational Partnerships Program. OERI funded six partnerships that included a focus on the transition from school to work. Most also focused on preparing students for a more technically sophisticated workplace than their parents experienced. Further, local program developers, community members, and business representatives expressed the belief (or hope) that if they developed a pool of highly trained workers, new industries would locate in the community. The "deconstructive" questions were:

1. Who participates in the training, by race, gender, and socioeconomic characteristics?

2. Do young people who participate in school-to-work programs get high tech jobs? Are there differences by race, gender, and socioeconomic group of those who get jobs?

3. Do new industries locate in the community? If so, did the existence of the labor pool influence the decision? Do locals get the "good" jobs?

The data are not yet in regarding either attraction of new industries or placements in high tech jobs. Students who participated in the school-to-work programs were more likely to get jobs than their classmates who did not, but the jobs tended to be part time, without benefits—characteristics of most new jobs across the nation. Two projects bucked this trend, both in the South.

One project leads to further questions about the distribution of benefits from the programs. That project, serving a prototypical rust belt city and its surrounding county, has many components. However, a major thrust is a school-to-work transition program, which includes a career education component for elementary and secondary schools. It also includes high tech-oriented training in conjunction with a major manufacturer at the regional vocational school that primarily serves students from the city. A third element in the transition component is training for careers related to robotics located at the suburban high school farthest from the city. Overall, there is a balance in the number of students served in the city and the county. A similar overall balance exists in the demographic characteristics of students. However, to the extent that data are available, it looks as if the most challenging, high tech programs have greater enrollment from the more middle class parts of the city and the suburbs.

The final example of deconstruction in evaluation also comes from the Educational Partnerships Program. In virtually every funded site, during the evaluation visit, someone would

say, "You know, it takes a whole village to raise a child." Deconstructing the village metaphor became important to the evaluation (Tushnet, 1994). In many instances, for example, the village was one that was—and should be—run by a small group of "elders," mainly white, mainly male. Their views on "raising a child" included helping students understand their place, creating "good workers," and, in general, perpetuating their privilege. In other instances, far fewer in number, the village metaphor was used in a way that called forth images of town meetings with broad participation and stimulating debate.

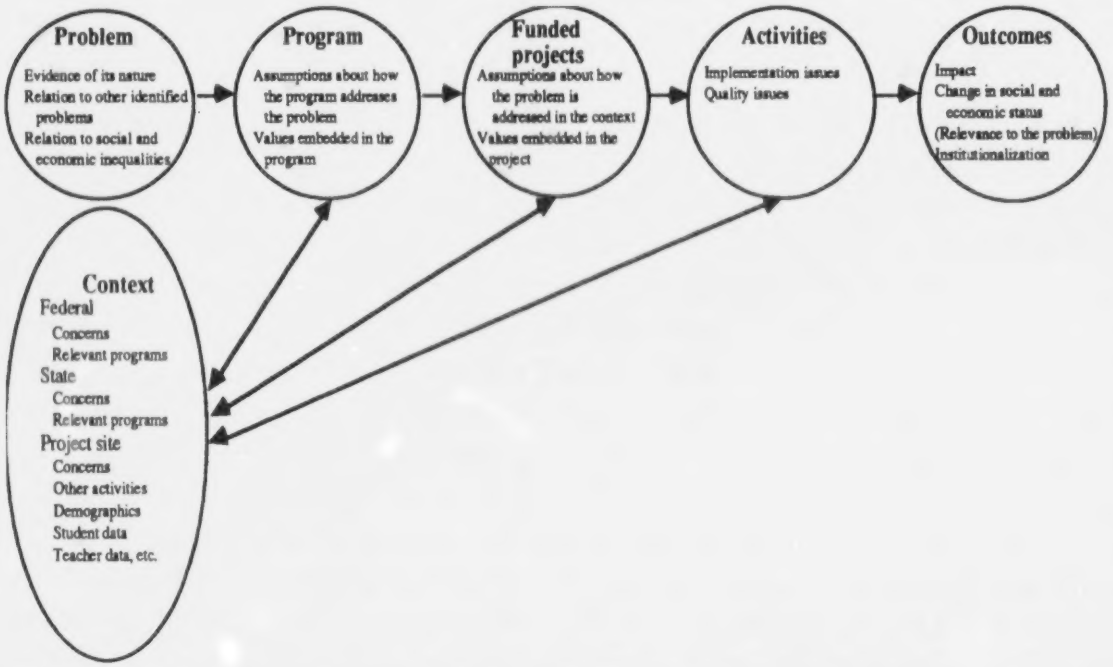
We pursued the metaphor by asking all participants, including those who received services from the partnerships, what meaning they took from its use. By doing so, we were able to develop multiple program narratives—giving voice to those frequently overlooked in studies. We found interesting contrasts in interpretations of the project, some of which were rooted in different understandings of the metaphor. Some projects, including most school-to-work partnerships, would have communicated more clearly, and experienced fewer problems, if they had relied on business-oriented metaphors, what we came to call the "language of the deal."

Participants in programs being evaluated are not always comfortable with deconstructive activities. From their perspective, evaluation is designed to help them manage the program, and raising questions about social impacts can be seen as inappropriate. In addition, deconstruction includes a view of social justice that is not necessarily shared by project staff. The view is rooted in Rawls' (1971) analysis of justice as fairness, and its role in evaluation has been well explicated by House (1993). The deconstructionist evaluator must take care to deconstruct his or her own questions and conclusions and provide opportunities for alternative views and interpretations to be included in reports.

## The Evaluation Framework

The framework for evaluation proposed in this paper is based on the theoretical stances described above. This section includes a discussion of the elements of the framework itself. Throughout, I indicate appropriate methods and how each element contributes to the usefulness of the approach. Figure 1 illustrates the evaluation framework for multisite program evaluation designs.

**Figure 1**
*Framework for the Development of Multisite Program Evaluation Designs*




### *Context*

Context, at the federal, state, and project site levels, frames what occurs during program implementation and has a major effect on program outcomes. On the federal and state levels, the concerns that led to the identification of the problem and other, related programs are key contextual matters. For project sites, context includes concerns, other activities, the history of innovation, demographics, and the characteristics of students, the community, and the professional staff of schools. Contextual variables relevant to an evaluation are identified by reference to the appropriate literature related to the content of the program and, particularly for demonstration programs, literature related to innovation and change. Consequently, other contextual factors also may be important for particular programs. For example, in the Star Schools Program, the existence or lack of an infrastructure to support technology was a key contextual feature influencing program implementation and outcomes. This section includes a discussion of sources of information about federal, state, and project site context.



BEST COPY AVAILABLE

**Federal and State Context**

At both the federal and state levels, key contextual information relates to the framing of particular policies. Studies of policy formation point to a number of key factors in development of new policies and their related programs. Pressures for change come from a number of sources, including social movements, issue networks (Heclo, 1978), and perceived crises (Polsby, 1984). According to Mazzoni (1991), these combine to create the context for policy change and, according to McDonnell & Elmore (1987), are combined with knowledge and assumptions about what types of actions would address the problem to determine the appropriate policy instrument. Further, as Wirt, Mitchell, and Marshall (1988) argue, how the concerns are responded to is largely framed by the state and federal "culture." State cultures differ, and federal cultures may change over time, but the key issue is the values by which decisions are made. For example, Wirt, Mitchell, and Marshall maintain that three values—choice, efficiency, and equity—frame much state education policy, and states balance them in different ways. In some states, and at times in the federal government, equity issues dominate. In others, and at other times at the federal level, choice is most important.

In addition to gathering information about the concerns that led to the development of the program, the evaluator gathers information about other concerns in the nation and states. This provides a broad framework for identifying the policy environment. Further, the evaluator also collects information about other, related activities being implemented.

The methods used to synthesize contextual information involve content analysis and interviews. Content analysis can be applied to various types of data, although it is most commonly thought of in relationship to documentary analysis. It allows information about phenomena of interest to be placed in standard categories, no matter what language is used. Content analysis enables quantification of textual data (Patton, 1980) that can be related to quantitative findings. The typology for coding information from policy and program documents is guided by preliminary understanding of the purpose and structure of the program under study (Hall, Hord, & Griffin, 1980; Rossman & Wilson, 1985). As analysis proceeds, new categories of interest are generated and findings from various data sources woven together (Schermerhorn, Williams, & Dickison, 1982). In complex, multisite program evaluations, content analysis is applied to the documents generating the program, including testimony, debate, and the Congressional Record.

Interviews are used to provide a basis for understanding the reasons the program took the form it did. The basic question is: "Why did you choose that approach?" (As the following sections reveal, interviews of program developers address context, definition of the problem, and program design.)

Context changes over time and may affect the design, implementation, and outcomes of the program. For example, the Star Schools Program always included goals related to integrating distance learning with educational reform. However, at the inception of the program, educational

reform was cast in terms of "more"—more advanced placement courses, more hours, more graduation requirements. This approach was reflected in the program's emphasis on full courses for high school students. Later, the educational reform was defined as "standards-based systemic reform," and Star Schools encompassed supplementary instruction that modeled the type of constructivist learning promulgated by the various educational standards projects. Evaluation reports that noted that changes in the program focus reflected changes in the policy context were enlightening to both federal program and local project staff. The substance of contextual change, particularly in the policy arena, is a prime candidate for deconstruction because such changes frequently reveal changes in the values held by key decisionmakers.

### Project Site

Evaluators are interested in concerns and activities at the project sites for the same reasons they are interested in federal and state concerns. That is, the concerns frame the program. Programs that address, either directly or indirectly, a number of highly salient issues are more likely to be successful than ones that are seen as tangential to the core issues at the site. For example, educational partnerships that focused on the transition from school to work were more likely to be successful in communities in which economic development was deemed important than were partnerships that focused on the gifted and talented.

Demographic data also are important, particularly for interpreting the impact of the program. Such data generally include information about students (gender, ethnicity, socioeconomic status), teachers (gender, ethnicity, years of experience), and the community. Data about student attendance, achievement, violence and vandalism, drug use, and language proficiency also are candidates for inclusion. Other variables also may be of interest, e.g., the number of computers in the schools or the history or interorganizational arrangements, depending on the focus of the program.

Evaluators use existing information, drawn from local, state, and federal databases, to develop the necessary information about the project site. However, there are limits to the amount of desired data available because states and districts frequently do not have sophisticated information systems. Further, few projects engage in the type of needs assessment that would build their information base. Consequently, evaluators gather as much relevant information as is available and paint a general picture of the context. If particular missing data take on importance during the course of the study, the evaluator must decide whether to alter the planned approach to data collection in order to gather additional contextual information. In the Star Schools Program, for example, the projects did not collect information about ethnic or socioeconomic characteristics of schools participating in the program. Without such data, it was impossible to tell if the program fulfilled its goal of equalizing educational opportunity. Consequently, evaluators sought such information.

The site context influences implementation and outcomes (Fullan, with Steigelbauer, 1991). Equally important, such data provide the basis for judging how equitable program effects are distributed.

### *Problem*

Most programs result from policymakers' and program developers' identification that a problem exists. Frequently, the problem has been an ongoing one, but is "noticed" in the policy arena as a result of key events. For example, public attention to sexual harassment, a phenomenon experienced by many women, was drawn as a result of the U.S. Supreme Court appointment of a man accused of workplace harassment. Similarly, most educational programs are developed in response to persistent problems in education that are brought to public consciousness at particular times.

Problems exist, but it is the definition of the nature of a problem that leads to the development of specific programmatic responses. The program theory and deconstructive approaches, then, are applied to how policymakers and program developers define the problem. This includes examination of the evidence used to establish the existence of the issue and justify the particular approach to addressing it embodied in the program. In the Star Schools Program, for example, information about the lack of advanced courses in rural areas was brought to bear as the program developed. The data were used politically by a Southern senator whose state was involved in major educational reform and who wanted federal support for the efforts, and substantively, in defining the curriculum subjects that were to be addressed.

The evaluator is equally concerned about relationships program developers see between the problem they are addressing and other problems. In the Star Schools Program, a concern for fostering educational use of technology was coupled with a general understanding that schools serving educationally disadvantaged children were likely to require assistance in purchasing equipment, which then could be used to support their learning. Perhaps more important, the assumed relationships are the basis for the program theory. Lipsey (1993) argues that program theory specifies the problem, including independent variables (generally, the treatment), and the nature of the subjects who are supposed to benefit from the treatment. Because program developers are seldom explicit about their theories, the evaluator begins by "discovering" the program theory.

The deconstructionist approach also aids understanding how program developers view relationships among problems. From that perspective, the evaluator probes for program developers' assumptions (or lack thereof) of how program will be distributed among participants. By deconstructing such presumed effects as well as each element in the theory, the evaluator gains understanding of intended, although not necessarily acknowledged, effects of the program on social relationships. For example, the Educational Partnerships Program generally reinforced

existing social relationships, as indicated by the examples of potential partners included in program documents. It included business and industry, institutions of higher education, social service agencies, and cultural institutions. The list lacked mention of unions, community-based organizations, or religious institutions—organizations that frequently have working and poor people in policymaking positions.

Evaluators use two methods to examine the evidence that a problem exists, how it is defined, and how it relates to other problems and social and economic inequalities: content analysis and interviews. Content analysis is implemented on program documents, grants announcements, guidelines to reviewers, and other relevant materials. Interviews are important for gaining understanding of the social meanings program developers bring to their task and how they interpret the problem. The next section describes the importance of interviews in describing the program theories, and the assumptions that frame program development.

### *Program*

Procedurally, evaluators give attention to the design of the program at the same time as they are seeking understanding of how program developers define the problem to which they are attending. The most important method at this stage is the use of interviews. Interviews have a special place in constructing and deconstructing program theories. They provide the opportunity for the evaluator and program developer to develop a shared understanding of the program. Through interviews, program developers reveal the social meanings (Denzin, 1994) they ascribe to program elements and are able to engage in critical or self-reflection (Kincheloe & McLaren, 1994). Although the evaluator enters the interview with a general protocol of questions to ask drawn from the content analysis, the purpose is to probe for the connections made, or not, by the program developer. It is frequently useful to pose hypothetical problems, or what I have told interviewees are "mischievous questions," to gain access to implicit theories. For example, from one perspective, educational partnerships can be viewed as extra or antidemocratic. That is, they give some decisionmaking authority over schooling to a group that has not been elected to make such decisions. Federal program developers at first seemed surprised at questions cast in this light but were able to explicate why they thought partnerships actually enhance democratic institutions through the building of community. Implicit theories, encompassing assumptions, and values thus became explicit.

The process of surfacing the assumptions and values that lay behind the program is what most program developers find useful. It provides an opportunity for them to reflect on what they have created, and federal policymakers, whether congressional staff or executive branch bureaucrats, believe they have little time in the ordinary course of business for such reflection. The process also raises their consciousness about contradictions in the program. They are then able to decide whether they wish to live with the contradictions in the interest of enhancing pluralism or

alter the program. For example, the Star Schools Program contains objectives related to three distinct goals: providing courses for students who would not otherwise have access to them; assisting in educational reform; and demonstrating educational applications of technology. Although these are not mutually exclusive goals, their existence has led to some confusion and much negative competition among projects. When federal program developers were asked about the three goals, they reflected a bit and acknowledged the differences among them for the first time. This enabled the program staff to communicate with projects in ways that decreased rancorous competition and helped build collaboration among the funded projects. Disagreements between staff at different projects remained, but the disagreements could be accepted more easily when projects did not fear that federal monitors preferred the objectives pursued by another project.

### *Funded Projects*

The same questions about assumptions are raised with regard to funded projects, which can be seen as the operationalization of program developers' aspirations. The attention to program theories and values held at each implementing site is a distinguishing mark of the approach to evaluation embraced here. That is, this approach recognizes that the same process of creating understanding of the problem and figuring out the activities that are likely to make things better occurs at each site as at the federal level. Differences among sites are *not* a matter of "policy drift" or "implementation failure" (Pressman & Wildavsky, 1973) but reflect the very nature of multisite programs. In part, differences in program theories reflect differences in the interpretation of the federal program. Goggin et al. (1990) say of federal legislation:

> Interpretation is a function of context...a single message...may be interpreted differently in different states. Even within a single state, the pluralistic and interactive nature of local-level policy adoption makes the probability of multiple, conflicting messages high." (p. 33)

The differences also reflect analysis of local conditions (context), including the form the problem takes and the knowledge, skills, and interests of those who will implement the program (Mowbray & Herman, 1991). Perhaps most important, however, the program theory at the federal level—the assumptions about relationships between elements of the "treatment" and how they will address the problem—is generally implicit. Consequently, project developers bring their own set of assumptions and values to bear on project design. In multisite programs, then, we can expect multiple program theories.

Evaluators address the same issues and use the same methods when they gather information from the funded projects as they do in addressing program design. They seek project staff members' interpretation of the problem, as revealed in the proposal and other relevant documents. As well, they seek to explicate the theory and values held by project developers about how their program addresses the problem in their particular context.

The information thus gained is useful in three ways. First, it is useful to the evaluators because it provides a framework for understanding implementation and outcomes. As such, it gives a map of where to look for activities and impacts, perhaps different from those indicated by federal policymakers. When there are differences between the guidance provided by the local project theory and the overarching program theory, evaluators are advised to seek information about both.

Second, analysis of the theories and values embedded in local project designs is illuminating to local project personnel in the same way(s) as the program analysis. Interviews serve to spur reflection. And, the summary analysis is shared with staff as a dialogue. Not all project managers greet deconstructive interpretations of their work with applause, but if the ground rules are that they are allowed to enter alternative interpretations, then most find the process helpful to program development at the site level. Indeed, the alternative interpretations also are indications of the theory and values that guide program implementation and also may help explain effects.

Third, the set of project analyses are useful to federal program developers. They can alter guidelines for projects if there are too many differences in theories or too few theories that address what the program developers think are the major issues. For example, the Educational Partnerships Program began with broad guidelines for eligible projects. Program managers assumed that most applicants for funding would be broad-based partnerships that addressed policy as well as program. The assumption was based on their reading of the literature and their disseminating research syntheses about partnerships to all potential applicants. In fact, however, most applicants had a small group of partners and focused on a fairly narrow change. After the first year of the evaluation, when OERI received project descriptions that included analyses of the local assumptions about the role of partnerships in school improvement, federal staff rewrote the grant guidelines to encourage more partnerships focused on systemic change. The guidelines became more explicit about the theory held by federal developers about how educational partnerships could contribute to educational reform.

### *Activities*

Program and project assumptions are played out through the activities implemented by the project. The activities operationalize the assumptions and values of project-level staff. It is sometimes useful to assess fidelity of implementation (Berman & McLaughlin, 1977), but it is generally more fruitful to study the activities that *are* implemented. The basic question is, "Who does what to whom and with what quality?"

The quality dimension addresses the extent to which project activities reflect what is known about the appropriate approaches to addressing the problem. Of course, definitions of quality reflect values, so the definitions are subject to deconstruction. For example, much current educational reform is designed to accomplish "standards-based systemic reform," and program

quality is judged by the extent to which activities supportive of such reform are implemented. But what are the "standards"? What social meanings do they embody? Apple (1990) provides a model of how deconstruction of standards can be done.

It is important to assess implementation for two reasons. First, evaluations that simply focus on outcomes may be evaluating "non-events" (Charters & Jones, 1973). Second, and more important, most new approaches comprise "innovation configurations" (Heck, Steigelbauer, Hall, & Loucks, 1981), with multiple activities making up the whole treatment. As a result of different program theories, each site is likely to configure the program somewhat differently. When the evaluation is completed, then, global statements about program impact are useless without an equal understanding of the relationship of particular configurations of activities to outcomes. In the Educational Partnerships Program, for example, a summary statement as to whether partnerships help support positive change would not provide information to those who wish to improve their partnerships or others seeking to start a new one. It is much more useful to conclude with statements like, "Educational partnerships that include *xyz* activities were more likely to have a positive impact on students than were those that did not. Further, community impact was gained when partners include *abc* agencies or individuals."

The methods for documenting and evaluating implementation include content analysis of documents, noting, for example, the number of workshops offered and the number of individuals attending. Interviews also are important sources of implementation data. Evaluators also can observe workshops, classes, or meetings of governing boards, using either qualitative protocols or quantifiable observation instruments. Surveys of project participants or recipients of services also are useful.

From the perspective espoused in this paper, data collection concerned with implementation must be fairly fine grained. It must attend to differential implementation for different participants. For example, there is much data to indicate that in classrooms, girls and boys are treated differently (Sadker & Sadker, 1994). Consequently, classroom observations of a "new" curriculum should include attention to gender differences during implementation. The deconstruction of implementation in the Star Schools Program led to the conclusion that white, rural, socioeconomically secure students were receiving advanced courses and minority, urban, poor students were receiving supplementary instruction. This understanding, then, was useful when results were interpreted.

The perspective on evaluation embraced in this paper includes data gathering activities from participants and recipients of program services. It is important that their voices be included in the study, although for students, particularly, issues of privacy must be dealt with. Despite the difficulties, evaluators should attend to those who are frequently voiceless. One educational partnership, for example, was a dropout "reclamation" alternative school, which provided education to students who had been out of school for over six months. Many were runaways or living on their own. The school provided self-esteem classes, computer-assisted individualized

instruction, workplace preparation, and opportunities to meet employers. Set in a mall, the school was small and had a low student-teacher ratio. The theory behind the program was explicit: The types of youth served by the program had low self-esteem, which stemmed from their many failures, and needed to develop self-discipline. When interviewed, students talked about the teachers' "compassion." For most, the school was a counseling center, and they internalized the values central to the self-esteem class, which were that they were responsible for their own actions and that their choices determined their success. Few were employed, and they ascribed their unemployment totally to their "choices" but were disappointed that school staff did not use their contacts to set up job interviews for them. Some were even angry. Both the positive and negative feelings speak to issues of implementation and quality.

The voice of service recipients can be heard through their actions as well as through interviews, although interviews are probably always useful in gaining understanding of the reasons for actions. In one Educational Partnerships project, for example, attendance at staff development sessions continuously decreased until, in the second year of the project, literally no one showed up. We found that the teachers did not see the connection of the staff development to the in-class activities and were quite articulate about its lack of relevance. Project staff were faced with a series of choices, including doing a better job of explaining the relationship; improving the relationship; or accepting that the in-class activities did not require staff development support. Their choice of the third alternative helped refine our understanding of their assumptions and values and supported a different interpretation of program outcomes from the first year's explanation.

Implementation evaluation is frequently cast as "process evaluation" and seen as formative. Indeed, evaluating the implementation of a program has an important role in formative evaluation. However, it also has a role in "impact evaluation" because it provides information necessary for determining if the treatment has merit (Scriven, 1991). From both vantage points, evaluations that include information about the implementation and quality of activities are judged useful by program and project developers. One function the evaluator can fill is to establish systems for monitoring implementation that can be used by local project staff so they can adjust activities as needed.

### *Outcomes*

The final category of concern is outcomes, which includes impact, including how effects are distributed among various groups, and institutionalization. Impact includes both the desired impacts of the program and projects and the unintended consequences. Institutionalization can be seen as the "stabilization of use of the innovation" (Huberman & Miles, 1986, p. 11) or "what's left" after external funding is no longer available. It involves the stabilization and continuation of practices and support for the program, with ongoing adjustments to changing context.

## Impact

The overarching impact question is: "Did the program 'fix' the problem it was designed to address?" Data about the extent to which it did can be drawn from a variety of sources, including surveys, student testing, interviews, and observations. The evaluators' decisions about appropriate measures are related to the desired outcomes. For example, it would be wrong to use standardized tests to measure the extent to which students engage in constructivist thinking. However, it might be appropriate to use standardized tests to see whether students taught in constructivist ways could meet "old" as well as "new" standards.

Frequently, evaluators seek data only from those upon whom the program is designed to have an impact. If a program is designed to improve student performance, for example, it is common to seek outcome data only from students. I argue that such information is not sufficient; it is necessary to find out the interpretations teachers or others give to the nature of the outcomes. Their views are important because they have ongoing contact with students. If they believe improvement in student performance is important and actually occurring, their interactions with students are likely to change. However, if they dismiss the change as trivial or doubt its existence, little change will occur.

Two examples drawn from a single school in an Educational Partnerships project illustrate the point. The substantive focus of the program was using local area networks (LANs) installed in participating inner-city schools as an impetus to changing teachers' instructional practices. The goal was to change from teacher-centered to "student-centered, inquiry-based" instruction. After two years in the program, student outcomes were strong on performance assessments. A first-grade teacher embraced the technology because she saw what it could do. She said, "I keep hearing about what first graders can do and not do. But now I know that first graders can do what first graders can do." Her students were engaged in demanding activities that required sophisticated critical thinking. In contrast, a fourth-grade teacher reading the same results said, "These performance tests don't count. When these kids go for the kind of jobs they'll be ready for, no one's going to care whether they can think." And her students spent most of their day doing typical "drill and skill" activities, interrupted occasionally by the use of the LAN.

## Institutionalization

Funders are particularly interested in finding out what happens when external support stops. One of the congressionally mandated questions asked about the Star Schools Program, for example, was, "How effective are telecommunications partnerships and programs after federal funding ceases?" Similarly, in the Educational Partnerships Program, OERI was interested in whether the activities espoused by the partnerships lasted after federal funding ceased and also whether the partnerships remained in existence. Unfortunately, funding for evaluations generally ceases at the same time as funding for the program so it is difficult to gather information about

institutionalization. However, indicators do exist, including whether funded positions are shifted to nonproject budgets and whether provisions are made to orient new staff members so they, too, can implement the program.

Data for evaluating the level of institutionalization come from content analysis of documents, including budgets, courses of studies, plans for staff development, and contracts for equipment maintenance. We found in the Star Schools Program that enrollment in courses declined each year without subsidies from the federal grant, and by the fourth year after federal funding ceased, they no longer existed.

Interviews with participants also are an important source of information about institutionalization, particularly when the study ends at the same time as the project. The questions focus on plans being made and values placed on the activities.

# CONCLUSION

I have argued that evaluations that address program developers' values and assumptions are particularly useful in multisite program evaluations. SWRL used this general approach in two national evaluations, and the results of the studies have been used by both project and program personnel. However, the analysis of the approach presented here is likely to raise questions about whether it is realistic. Can any evaluation support all the activities delineated? In this concluding section, I address that concern.

Clearly, the approach requires evaluators with great skill and knowledge. They must have the appropriate technical training to: design valid interview protocols, observational schemes, and survey instruments; develop sampling plans; conduct content analysis; and analyze and interpret data. More important, they must have substantive knowledge of the program focus in order to recognize assumed relationships that are part of the program theories espoused by program and project developers. And most difficult, they must have a broad and deep understanding of the society in order to engage in the deconstructive tasks.

Does such a person exist? Probably not, particularly because evaluators generally work in a number of content areas. Even those of us who confine ourselves to education evaluate a variety of different substantive programs, including those designed to improve curriculum (and in different substantive areas), school organization, and the achievement of particular groups of students. However, it is possible to gather teams of people who can share their knowledge while holding particular responsibility for one aspect of the evaluation.

The second practical barrier to implementing the evaluation approach presented here lies in the amount of resources it requires. As outlined above, the evaluation approach is labor intensive and requires multiple types of data from multiple sources. Seldom are sufficient funds available. However, the approach offered in this paper is a model, and I do not expect any agency, including SWRL, will ever fulfill all its requirements.

The approach has shown itself to be a useful model in our work. It enables us to create meaningful frameworks for each evaluation we conduct. We are aware of what we lose when we decide not to focus on a particular part of the model, and can open conversations with clients about the tradeoffs involved in our design. Perhaps most important, the approach is a constant reminder to probe beneath the surface to the social meanings of program elements and effects. It was natural for SWRL evaluations to deconstruct the different meanings of the village metaphor in the Educational Partnerships Program and to seek the data that reveal the differences in opportunities for students in the Star Schools Program. In both cases, program developers valued what they saw as unusual "insights."

Even when not fully used, the model creates a positive environment for evaluation use. When we put together the program theories, for example, we share them with key audiences. The sharing opens a dialogue in which program staff can say, "That isn't what I think. This is what I

think about why the activities I'm putting in place will solve the problem." The explication helps program developers in their reflections about their program. Their reactions help the evaluators refine the evaluation design in order to gather more appropriate information. The interactions, then, build a basis for evaluation use.

In sum, the evaluation approach focuses on the values and assumptions that program and project developers use in designing how they will address a problem. Because many of the assumptions are implicit, the approach is immediately useful. Further, the approach recognizes that project developers may hold different program theories from one another and from the program developers. By making the differences clear, the evaluators decrease competition among developers, which may lead to greater collaboration. The clarity also enables program developers to refine their announcements and/or selection criteria and decide to provide technical assistance. The deconstructive mode advances attention to differential effects of a program

# REFERENCES


Apple, M. W. (1990). *Ideology and curriculum* (2nd ed.). New York: Routledge & Kegan Paul.

Berger, P., & Luckmann, T. (1966). *The social construction of reality*. Garden City, NY: Doubleday.

Berman, P., & McLaughlin, M. W. (1977). *Federal programs supporting educational change: Vol. VIII. Implementing and sustaining innovations*. Santa Monica, CA: Rand Corporation.

Charters, W., & Jones, J. (1973). On the risk of appraising non-events in program evaluation. *Educational Researcher, 2*(11), 105-110.

Chen, H.-T. (1989). The theory-driven perspective. *Evaluation and Program Planning, 12*(4), 297-312.

Chen, H.-T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283-302.

Denzin, N. K. (1992). *Symbolic interactionism and cultural studies*. Cambridge, UK: Basil Blackwell.

Denzin, N. K. (1994). The art and politics of interpretation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 500-515). Thousand Oaks, CA: Sage.

Derrida, J. (1976). *Of grammatology* (G. C. Spivak, Trans.). Baltimore: Johns Hopkins University Press.

Derrida, J. (1981). *Positions* (A. Bass, Trans.). Chicago: University of Chicago Press.

Education Acts, Amendments, 1991, §20, III U.S.C. §4081 Pub. L. No. 102-103 (Aug. 17, 1991).

Flax, J. (1987). Postmodernism and gender relations in feminist theory. *Signs, 14*, 621-643.

Fullan, M. G., with Stiegelbauer, S. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Goggin, M. L., Bowman, A. O'M., Lester, J. P., & O'Toole, Jr., L. J. (1990). *Implementation theory and practice: Toward a third generation*. Boston: Harper Collins.

Hall, G. E., Hord, S., & Griffin, G. (1980, April). *Implementation at the school building level: The development and analysis of nine mini-case studies*. Paper presented at the annual meeting of the American Educational Research Association.

Heck, S., Steigelbauer, S., Hall, G. E., & Loucks, S. (1981). *Measuring innovation configurations: Procedures and applications*. Research and Development Center for Teacher Education. Austin: University of Texas.

Heclo, H. (1978). Issue networks and the executive establishment. In A. King (Ed.), *The new American political system* (pp. 87-124). Washington, DC: The American Enterprise Institute.

House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage.

Huberman, A. M., & Miles, M. B. (1986). *Innovation up close*. New York: Plenum.

Kincheloe, J. L., & McLaren, P. L. (1994). Rethinking critical theory and qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 138-157). Thousand Oaks, CA: Sage.

Lipsey, M. W. (1993, Spring). Theory as method: Small theories of treatments. In L. B. Sechrest & A. G. Scott (Eds.), *New directions for program evaluation, 57*, 5-38. San Francisco: Jossey-Bass.

Mazzoni, T. L. (1991). Analyzing state school policymaking: An arena model. *Educational evaluation and policy analysis, 13*(2), 115-138.

McDonnell, L. M., & Elmore, R. F. (1987). Getting the job done: Alternative policy instruments. *Educational evaluation and policy analysis, 9*(2), 133-152.

Mowbray, C. T., & Herman, S. E. (1991, Summer). Using multiple sites in mental health evaluations. In R. S. Turpin & J. M. Sinacore (Eds.), *New directions for program evaluation, No. 50*. San Francisco: Jossey-Bass.

Olesen, V. (1994). Feminisms and models of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 158-174). Thousand Oaks, CA: Sage.

Patton, M. (1980). *Qualitative evaluation methods*. Beverly Hills: Sage.

Polsby, N. (1984). *Political innovation in America*. New Haven, CT: Yale University Press.

Pressman, J. L., & Wildavsky, A. (1973). *Implementation*. Berkeley, CA: University of California Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Belknap.

Rossman, G. B., & Wilson, B. L. (1984). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review, 9*(5), 627-643.

Sadker, M., & Sadker, D. (1994). *Failing at fairness: How America's schools cheat girls*. New York: Charles Scribner's Sons.

Schermerhorn, L. L., Williams, L., & Dickison, A. (Eds.). (1982). *Project COMPAS: A design for change*. East Peoria, IL: Illinois Central College (ERIC Document Reproduction Service ED 219100).

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Schwandt, T. A. (1994). Contructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118-137). Thousand Oaks, CA: Sage.

Sinacore, J. M., & Turpin, R. S. (1991, Summer). Multiple sites in evaluation research: A survey of organizational and methodological issues. In R. S. Turpin & J. M. Sinacore (Eds.), *New directions in program evaluation, 50*, 5-18. San Francisco: Jossey-Bass.

Smith, N. L. (1993, November). *Contingencies and investigative variations of program theory-driven evaluations*. Paper presented at the annual meeting of the American Evaluation Association, Dallas.

Tushnet, N. C. (1993, November). *The problem of evaluation use with multiple audiences*. Paper presented at the annual meeting of the American Evaluation Association, Dallas.

Tushnet, N. C. (1994, April). *Does it really take a whole village to raise a child: Mixing metaphor and meaning in the Educational Partnerships Program.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Tushnet, N. C., Bodinger-deUriarte, C., vanBroekhuizen, L. D., Manuel, D. M., Danzberger, J., & Clark, M. (1993). *Documentation and evaluation of the Educational Partnerships Program: Year 2 report.* Los Alamitos, CA: Southwest Regional Laboratory.

Tushnet, N. C., Schwager, M. T., Fleming-McCormick, T., Manuel, D. M., Nyre, G. F., Millsap, M. A., Chase, A. M., Johnston, K. S., Johnson, S. L., Beckford, I., McPherson, A., Levin, M., & Brigham, N. (1994). *The Star Schools Program: Final report.* Los Alamitos, CA: Southwest Regional Laboratory.

Vidich, A. J., & Lyman, S. M. (1994). Qualitative methods: Their history in sociology and anthropology. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 23-59). Thousand Oaks, CA: Sage.

Wirt, F., Mitchell, D., & Marshall, C. (1988). Culture and educational policy: Analyzing values in state policy systems. *Educational Evaluation and Policy Analysis, 10*(4), 271-284.